GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Minor / Hons. Program

Program Level: UG

Branch: Minor / Hons. Data Science

Subject Code: BE050AT011

Subject Name: Introduction to AI and ML with Data Science

w. e. f. Academic Year:	2026-27
Semester:	5
Category of the Course:	Core Courses

Prerequisite:	Basic knowledge of Mathematics, basic calculus, and probability fundamentals, basic understanding of Python programming, Statistics (Mean, variance, correlation, and data interpretation), Data Handling Concepts: Basics of data representation, datasets, and simple data analysis.
Rationale:	This course provides a strong foundation for advanced learning in data science, analytics, artificial intelligence, and intelligent systems, preparing them for industry-oriented roles and higher academic pursuits.

Course Outcome:

After Completion of the Course, Student will able to:

Sr. No.	CO statement	Marks % Weightage
CO-1	Comprehend the fundamental principles, techniques, and workflow of Machine Learning (ML) as applied in Data Science	10%
CO-2	Understand various categories of ML models, including regression, classification, clustering, and dimensionality reduction and Understand Artificial Intelligence concept.	25%
CO-3	Apply suitable Machine Learning algorithms such as Linear/Logistic Regression, K-Means, SVM, and PCA to solve real-world data-driven problems using Python.	25%
CO-4	Analyse and interpret model performance using metrics like accuracy, precision, recall, F1-score, and confusion matrix.	20%
CO-5	Compare and select optimal models using advanced techniques like Ridge and Lasso regression, cross-validation, and bias-variance trade off.	20%

Teaching and Examination Scheme:

Teaching Scheme (in Hours)					Total Credits= TH/30	Assessment Pattern and Marks					Total Marks
						Theory		Tutorial / Practical			
L	T	PR	PBL *	TH	C	ESE (E)	PA (M)	PA (I)	PBL (I)	ESE (V)	
45	30	00	45	120	4	70	0	0	30	50	150

Course Content:

Unit No.	Content	No. of Hours	% of Weightage

1.	Linear Regression Definition and need of Regression in Data Science. Concept of correlation and least squares method. Linear Regression equation and line fitting. Evaluation metrics: MAE, MSE, RMSE, R^2. Implementation of Linear Regression using Python (NumPy, scikit-learn). Case Study: Predict housing prices using dataset from Kaggle.	6	20 %
2.	Logistic Regression Sigmoid function and decision boundary. Binary and Multiclass Logistic Regression. Evaluation using Confusion Matrix, Precision, Recall, F1-score. Python implementation using scikit-learn. Case Study: Predict employee attrition or customer churn.	8	20%
3.	Basics of Artificial Intelligence Introduction to Artificial Intelligence Definition and scope of AI History and evolution of AI Applications of AI (healthcare, robotics, finance, etc.) Difference between AI, Machine Learning, and Deep Learning Types of AI: Weak AI, Strong AI, Super AI	4	20%
4.	Clustering and PCA Definition of clustering and unsupervised learning. K-Means and Hierarchical Clustering concepts. Principle Component Analysis (PCA): Dimensionality reduction. Applications: Market segmentation, recommendation systems. Case Study: Customer segmentation using K-Means and PCA visualization.	9	20%
5.	Support Vector Machine (SVM) Concept of hyperplane, margin and kernel functions. Linear and Non-linear SVM. Applications: text classification, spam detection, alphabet recognition. Implementation using scikit-learn. Case Study: Handwritten alphabet recognition using SVM.	9	20%

	Total	45	100%

Suggested Specification Table with Marks (Theory):

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
7	**17**	**33**	**21**	**18**	**4**

Where R: Remember; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create (as per Revised Bloom's Taxonomy)

References/Suggested Learning Resources:
(a) Books:

1. Machine Learning using Python , U Dinesh Kumar and Manaranjan Pradhan, John Wiley & Sons.
2. Advanced Data Analytics Using Python: With Machine Learning, Deep Learning by By Sayan ukhopadhyay, Apress.
3. Practical Data Mining" by Monte F. Hancock, Auerbach Publication.
4. "Machine Learning for Absolute Beginners: A Plain English Introduction (Second Edition)" by Oliver Theobald.
5. Practical Data Science with R, Nina Zumel, John Wiley & Sons.
6. Python for Data Science for Dummies, John Paul Mueller, Luca Massaron, John Wiley & Sons.
7. Big Data and Analytics, Seema Acharya and Subhashini Chellappan, Wiley Publication.

(b) Open source software and website:

1. **Python (Anaconda Distribution)** – Open-source programming environment for Data Science and Machine Learning (https://www.anaconda.com/products/distribution)
2. **Jupyter Notebook / JupyterLab** – Interactive notebook for Python programming, data visualization, and analysis (https://jupyter.org/)
3. **Google Colab** – Cloud-based Python and Jupyter environment for Data Science and Machine Learning practice (https://colab.research.google.com/)
4. **Kaggle** – Open platform for datasets, competitions, and collaborative Data Science notebooks (https://www.kaggle.com/)

Course Practical List:

Sr. No	Practical	CO

Practical Set-1:(Advance of Python)		**CO1,CO2**
1.1	Implement a program for defining, optimizing, and evaluating mathematical expressions involving multi-dimensional arrays with use of Theano Libraries for Symbolic Computation.	
1.2	Implement a program for addition two constant value with use of Tensorflow Libraries.	
1.3	Implement a program for compute gradient of y = x^2 with use of PyTorch Libraries	
Practical Set-2: (Linear Regression)		**CO1,CO2**
2.1	Implement a program for Simple Linear Regression using one variable.	
2.2	Implement a program Multiple Linear Regression	
2.3	Implement a program Linear Regression using dataset	
2.4	Implement a program Polynomial Regression	
Practical Set-3: (Logistic Regression)		**CO2,CO3**
3.1	Implement a program Logistic Regression (Binary Classification)	
3.2	Implement a program Logistic Regression on Real Dataset (CSV)	
3.3	Implement a program Evaluate Model Accuracy	
Practical Set-4: (Basics of AI)		**CO2,CO3**
4.1	Implement a program Concepts of Prolog like Fact Rules & Query.	
4.2	Implement a program Comparison Operators, Arithmetic operator solve by using Prolog.	
Practical Set-5: (k-Mean Clustering)		**CO2,CO3**
5.1	Implement a program K-Means Clustering (Basic Example).	
5.2	Implement a program Find Optimal Number of Clusters using the Elbow Method.	
Practical Set-6: (Hierarchical & DBScan Clustering)		**CO2,CO3**
6.1	Implement a program Hierarchical Agglomerative Clustering.	
6.2	Implement a program DBSCAN Clustering.	
6.3	Implement a program Customer Segmentation using Clustering.	
Practical Set-7: (PCA)		**CO3**
7.1	Implement a program PCA on Simple Dataset	
7.2	Implement a program Plot Variance Explained by Each Principal Component	
7.3	Implement a program Compare Model Accuracy Before and After PCA	
Practical Set-8: (Basics of SVM)		**CO3,CO4**
8.1	Implement a program Basic SVM Classifier on Synthetic Data	
8.2	Implement a program Visualize SVM Decision Boundary (2D Data	
8.3	Implement a program Evaluate SVM Model Performance (Compute accuracy, confusion matrix, precision, recall, and F1-score)	

Practical Set-9: (Advance of SVM)		**CO4,CO5**
9.1	SVM for Spam Email Detection	
9.2	SVM for Handwritten Alphabet / Digit Recognition	
Practical Set-10: (Lasso & Ridge Regression)		**CO4,CO5**
10.1	Implement Lasso Regression on Simple Data	
10.2	Implement a program Ridge Regression on Simple Data	

Suggested Activities for Students: If any

Suggested Activities for Students: Mini Project (45 Hours) Mini Project Design Instructions (Develop a Python-based mini-project integrating concepts from all practical sets.)

* * * * * * *